

A Hybrid Sunshine–Temperature–Humidity Model for Estimating the Monthly Mean Daily Global Solar Radiation of Tébessa, Algeria

Tidjani Boutora

boutora.tidjani@gmail.com

Mohamed Chérif Messaadia University of Souk Ahras

Article

Keywords: solar energy, global solar radiation, relative sunshine duration, semi-empirical modeling, clearness index, multicollinearity, semi-arid climate, Tébessa, Algeria

Posted Date: September 22nd, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10786489/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Reliable knowledge of the global solar radiation reaching a horizontal surface is a prerequisite for the design of any solar conversion system, yet pyranometric stations remain scarce over the Algerian territory. Semi-empirical correlations driven by routinely recorded meteorological variables therefore remain the practical alternative. In a previous work the authors proposed a two-variable model, $H/H_0 = aT^2 + bRH + c$, calibrated for Souk Ahras; when transferred to continental sites its relative error rose to 6.82% at Tébessa and 8.93% at Sétif, which suggested that thermal predictors alone do not describe atmospheric transmissivity in semi-arid highlands. The present study tests that hypothesis by adding the relative sunshine duration $\sigma = S/S_0$ as a third regressor and calibrating the resulting four-parameter correlation for Tébessa (35.404°N, 8.124°E, 850 m a.s.l., BSk climate) on the monthly database of the Renewable Energy Development Centre completed by the sunshine record of the Algerian National Meteorological Office. Six candidate correlations were fitted and assessed by leave-one-month-out cross-validation against ten statistical indicators. The proposed correlation, $H/H_0 = 0.14900 \sigma + 0.000137 T^2 + 0.15222 RH + 0.38402$, gives $RMSE = 0.243 \text{ MJ}\cdot\text{m}^{-2}\cdot\text{day}^{-1}$ ($nRMSE = 1.31\%$), $MBE = +0.0032 \text{ MJ}\cdot\text{m}^{-2}\cdot\text{day}^{-1}$ and $R^2 = 0.9984$, against $RMSE = 0.254 \text{ MJ}\cdot\text{m}^{-2}\cdot\text{day}^{-1}$ and $MBE = +0.0072 \text{ MJ}\cdot\text{m}^{-2}\cdot\text{day}^{-1}$ for the previous formulation: the systematic bias is reduced by 56% and the Student statistic falls from 0.094 to 0.043, but the root mean square error improves by only 4.2%. The reason is quantified rather than assumed: on a monthly climatological normal the three predictors are almost perfectly collinear (variance inflation factors of 181 for σ , 633 for T^2 and 142 for RH ; correlation of -0.96 between σ and RH), so that the sunshine coefficient, although carrying the largest sensitivity of the three ($\pm 1.78\%$ of K_T for a $\pm 10\%$ perturbation), is not individually significant at the 5% level ($p = 0.234$). The practical conclusion is that the sunshine term improves the accuracy and the bias of the estimate and should be retained, but that a decisive separation of the three predictors requires daily rather than monthly-normal calibration data.

1. INTRODUCTION

The Algerian energy balance is still almost entirely built on hydrocarbons, and the national economy remains structurally dependent on their export. The volatility of oil prices over the last decade, together with the commitments taken by the international community for the protection of the climate, has made the diversification of the national energy mix an economic necessity rather than an environmental option. The Algerian government has consequently launched a series of programmes aimed at the progressive substitution of fossil resources by renewable ones, solar energy occupying the first rank given the size of the resource available over the national territory.

Any solar project, however, begins with the same question: how much energy actually reaches the ground at the site considered? The answer requires long and continuous series of measured global solar radiation, and this is exactly where the difficulty lies. Pyranometric stations are expensive to purchase, to install and above all to maintain; as a result the Algerian radiometric network covers only a handful of sites, and several of the wilayas with the highest solar potential possess no radiation record at all. The

situation is far from being specific to Algeria: it is the rule rather than the exception in developing countries, and it explains why the estimation of global solar radiation from ordinary meteorological variables has remained an active field of research for almost a century.

Two large families of correlations have historically dominated the literature. The first, initiated by Ångström and later reformulated by Prescott, relates the clearness index to the relative sunshine duration and remains the reference against which any new proposal is measured. The second family replaces the sunshine record – which is itself not always available – by thermal descriptors such as the mean air temperature, the diurnal temperature range or the relative humidity. A third, more recent family combines several of these predictors in hybrid formulations, and a fourth relies on machine-learning architectures trained on large data sets. The abundance of published correlations is not a sign of redundancy but of a well-known limitation: the regression coefficients are site-dependent, because they absorb the local aerosol load, the cloud regime and the topographic context of the station at which they were fitted.

The present work starts from a transfer failure observed in the authors' own previous study [1]. In that work a semi-empirical model of the form $H/H_0 = aT^2 + bRH + c$ was calibrated for Souk Ahras and then applied to ten Algerian cities. The agreement was excellent over Mostaganem (2.49%) and Adrar (4.71%), but clearly degraded over Sétif (8.93%), Tamanrasset (7.19%) and Tébessa (6.82%). The three sites at which the model performed worst share a marked continentality and an elevation exceeding 800 m. The working hypothesis of the present paper is that this pattern reflects the absence, among the predictors, of any direct descriptor of atmospheric transmissivity: air temperature and relative humidity respond to the incoming radiation rather than determining it, and their response is both delayed and damped by the thermal inertia of the boundary layer.

The variable that measures transmissivity directly is the sunshine duration. By definition, the number of hours during which the direct beam exceeds the World Meteorological Organization threshold of $120 \text{ W}\cdot\text{m}^{-2}$ is a measure of the time during which the atmosphere is transparent to the beam component. This is why sunshine-based correlations have consistently outperformed purely thermal ones, a conclusion recently reconfirmed at continental scale by Lan et al. [2] over the climate zones of China and by Banihashemi Dehkordi et al. [3] over six Iranian stations in arid and semi-arid climates.

The objection raised in [1] against the use of sunshine data – that such measurements are not available in developing countries – was legitimate at the time of writing but has since been overtaken by the evolution of the observation systems. The Algerian National Meteorological Office operates a network of more than fifty stations recording sunshine hours, whose spatial value has been demonstrated by Benmouiza [4] through kriging; and the third edition of the Surface Solar Radiation Data Set – Heliosat released by the EUMETSAT Satellite Application Facility on Climate Monitoring now delivers a sunshine duration product at $0.05^\circ \times 0.05^\circ$ covering the whole of Africa from 1983 onwards [5].

The objective of this paper is therefore to extend the correlation of [1] by introducing the relative sunshine duration as a fourth degree of freedom, and to calibrate the resulting model for the wilaya of

Tébessa, a semi-arid highland site where the previous formulation showed one of its largest deviations. The contributions are threefold: (i) a four-parameter hybrid correlation calibrated and cross-validated for a BSk climate; (ii) a quantitative assessment, based on ten statistical indicators and a leave-one-month-out procedure, of how much the sunshine term actually contributes once the thermal predictors are present; and (iii) an explicit diagnosis of the multicollinearity that limits that contribution when the calibration set is a monthly climatological normal, together with the recommendation that follows from it.

The remainder of the paper is organised as follows. Section 2 describes the study area and the databases. Section 3 reviews the existing correlations and organises them into four families. Section 4 recalls the astronomical background. Section 5 introduces the proposed model and the calibration strategy. Section 6 presents the results, Section 7 discusses them, and Section 8 concludes.

2. STUDY AREA AND DATA SOURCES

2.1 Description of the site

The wilaya of Tébessa occupies the extreme north-east of Algeria, along the Tunisian border, at the junction between the eastern Saharan Atlas and the high plains of Constantine. The city lies at 35° 24' 15" N and 08° 07' 27" E, at an elevation of about 850 m above sea level; the synoptic station of Tébessa – Cheikh Larbi Tébessi, from which the meteorological records used here originate, is situated at 811 m.

According to the Köppen–Geiger classification the site belongs to the BSk type, a cold semi-arid climate. This is a genuinely different context from that of Souk Ahras, classified Csa and lying at 653 m only 100 km further north. Table 1 summarises the geographical and climatic parameters obtained from the record analysed in this work.

Table 1
Geographical and climatic characteristics of the study site

Parameter	Value	Unit / remark
Site	Tébessa (Cheikh Larbi Tébessi station)	—
Latitude φ	35.404 N	decimal degrees; used in Eqs. (1)–(4)
Longitude	8.124 E	decimal degrees; not used in the computation of H_0
Elevation	850 (city) / 811 (station)	m a.s.l.
Köppen–Geiger class	BSk	cold semi-arid
Annual mean air temperature	16.59	°C (7.20 in January, 27.32 in July)
Mean diurnal range ΔT	12.48	°C (9.59 in December, 15.85 in July)
Annual mean relative humidity	52.7	% (70.1 in December, 30.8 in July)
Annual mean sunshine duration	8.87	$\text{h}\cdot\text{day}^{-1}$, i.e. about $3240 \text{ h}\cdot\text{year}^{-1}$
Annual mean relative sunshine duration σ	0.732	– (0.659 in December, 0.853 in July)
Annual mean global irradiation H	18.58	$\text{MJ}\cdot\text{m}^{-2}\cdot\text{day}^{-1}$ (9.92 in December, 27.01 in July)
Annual mean clearness index K_T	0.618	– (0.591 in January, 0.664 in July)

Two arguments justify the choice of this site. The first is scientific: Tébessa is one of the three locations at which the previous model [1] showed its largest deviation, so the site constitutes a documented test case rather than an arbitrary one. The second is practical: the wilaya combines an intense mining and cement industry, hence a substantial and growing electricity demand, with large expanses of arid land suitable for photovoltaic plants.

The values of Table 1 are themselves informative. A clearness index between 0.59 and 0.66 and a relative sunshine duration between 0.66 and 0.85 place Tébessa among the well-insolated but not hyper-arid sites: the sky is clear enough for the annual variation of the clearness index to be modest, which, as Section 7 shows, has direct consequences for the identifiability of the model coefficients.

2.2 Data sources

The primary database used in this study is that of the Renewable Energy Development Centre (CDER), which constitutes the national reference for solar resource assessment in Algeria and which was already the validation source of the previous work [1]. The CDER record provides the monthly mean daily global

solar irradiation on a horizontal surface. Because the proposed model requires a sunshine input that the CDER database does not contain, the radiation record is completed by the sunshine duration series of the Algerian National Meteorological Office (ONM) for the Tébessa station, together with the corresponding air temperature and relative humidity series. Table 2 lists the databases and their role.

Table 2
Databases used in this study

Database	Variables retrieved	Resolution and coverage	Role
CDER – Renewable Energy Development Centre	Monthly mean daily global irradiation H on a horizontal surface	Monthly; national coverage	Reference values of H
ONM – Algerian National Meteorological Office	Sunshine duration S; air temperature T, Tmax, Tmin; relative humidity RH	Monthly, Tébessa station	Meteorological predictors
CM SAF SARAH-3 (EUMETSAT) [5]	Sunshine duration SDU; surface incoming shortwave SIS	0.05° × 0.05°; 1983 onwards	Recommended for the extension of the model to station-free sites (Section 7)

2.3 Nature of the record and its statistical consequence

The record supplied for Tébessa covers the ten years 2015–2024, that is, 120 monthly values. Its internal structure must, however, be stated explicitly, because it governs the whole of the statistical treatment that follows. The inter-annual standard deviation of the monthly global irradiation amounts to 0.31% of its mean, and that of the clearness index to less than 0.5%; the ten years are therefore very nearly identical to one another. What the database provides is a monthly climatological normal rather than a genuine time series – which is what a national solar atlas is designed to deliver, and is entirely legitimate – but it means that the effective sample size is twelve and not one hundred and twenty.

Two consequences follow. First, a calibration/validation split by year would be meaningless, since the validation years would be almost exact copies of the calibration years; the models are therefore assessed by leave-one-month-out cross-validation, in which each monthly value is predicted by a correlation fitted on the eleven remaining months. Second, all the confidence intervals and significance tests reported below are computed on twelve degrees of freedom, which is the honest count. Reporting them on 120 would inflate every Student statistic by a factor of about three and would give a false impression of precision.

Records were screened before use: months with an impossible relative sunshine duration (σ 1), a clearness index outside the interval [0.35, 0.80] or a minimum temperature exceeding the corresponding maximum were to be rejected. No value of the final record failed these tests.

3. REVIEW OF EXISTING GLOBAL RADIATION MODELS

The correlations proposed since 1924 are numerous, and a purely chronological listing does not by itself reveal the logic of their evolution. They are therefore grouped here into four families according to the nature of the information they exploit. Family A gathers the sunshine-based correlations, which relate the clearness index to the relative sunshine duration through a linear, polynomial or logarithmic function. Family B gathers the temperature-based correlations, in which the diurnal temperature range or the mean temperature replaces the sunshine record. Family C gathers the hybrid correlations, which combine sunshine, temperature, humidity and sometimes altitude or latitude. Family D gathers the data-driven approaches. Table 3 summarises the most representative members of the first three families, to which the proposed model belongs.

Table 3
Representative correlations for the estimation of the monthly mean daily global solar radiation

No.	Author (year)	Analytical form	Family
01	Ångström (1924) [6]	$H/H_0 = a + b (S/S_0)$	A
02	Glover and McCulloch (1958) [7]	$H/H_0 = 0.29 \cos \varphi + 0.52 (S/S_0)$	A
03	Page (1961) [8]	$H/H_0 = 0.23 + 0.48 (S/S_0)$	A
04	Swartman and Ogunlade (1967) [9]	$H/H_0 = a + b (S/S_0) + c RH$	C
05	Iqbal (1979) [10]	$H_d/H = 0.791 - 0.635 (S/S_0)$	A
06	Khogali et al. (1983) [11]	$H/H_0 = 0.191 + 0.571 (S/S_0)$	A
07	Ögelman et al. (1984) [12]	$H/H_0 = a + b (S/S_0) + c (S/S_0)^2$	A
08	Bahel et al. (1987) [13]	cubic polynomial in (S/S_0)	A
09	Gopinathan (1988) [14]	$H/H_0 = f(\varphi, Z, S/S_0)$	C
10	Newland (1989) [15]	$H/H_0 = 0.34 + 0.40 (S/S_0) + 0.17 \log (S/S_0)$	A
11	Akinoğlu and Ecevit (1990) [16]	$H/H_0 = 0.145 + 0.845 (S/S_0) - 0.280 (S/S_0)^2$	A
12	Samuel (1991) [17]	cubic polynomial in (S/S_0)	A
13	Aksoy (1997) [18]	$H/H_0 = 0.148 + 0.668 (S/S_0) - 0.079 (S/S_0)^2$	A
14	Ampratwum and Dorvlo (1999) [19]	$H/H_0 = 0.6376 + 0.2490 \log (S/S_0)$	A
15	Chegaar and Chibani (2001) [20]	$H/H_0 = a + b (S/S_0)$, coefficients per Algerian site	A
16	Chen et al. (2004) [21]	$H/H_0 = a (T_{max} - T_{min})^{0.5} + b$	B
17	Bakirci (2009) [22]	$H_c/H_0 = a + b (S/S_0) + c (S/S_0)^2$	A
18	Olayinka (2011) [23]	$H/H_0 = a_1 + a_2 T_{max} + a_3 \sigma$	C
19	Adaramola (2012) [24]	$H/H_0 = a_1 + a_2 T_m$; $H/H_0 = a_1 + a_2 (T_{min}/T_{max})$	B
20	Adhikari et al. (2013) [25]	$H/H_0 = a + b (S/S_0) + c (T_{avg}/T_{max}) + d \ln(RH)$	C
21	Li et al. (2013) [26]	$H/H_0 = a_1 (1 + a_2 RH) \Delta T$	C
22	Mecibah et al. (2014) [27]	best-model selection over eight Algerian sites	A
23	Yaiche et al. (2014) [28]	sunshine-based mapping of the Algerian territory	A
24	Hassan et al. (2016) [29]	$H/H_0 = a_1 + a_2 \Delta T^{0.5}$	B
25	Jahani et al. (2017) [30]	cubic polynomial in ΔT	B

No.	Author (year)	Analytical form	Family
26	Fan et al. (2018) [31]	$H/H_0 = a_1 + a_2 \Delta T + a_3 \Delta T^{0.25} + a_4 \Delta T^{0.5}$	B
27	Ekici and Teke (2018) [32]	Buckingham- π formulation in RH, T _{min} and ΔT	C
28	Yıldırım et al. (2018) [33]	$H/H_0 = a_1 + a_2 RH + a_3 \sigma + a_4 \sigma^2 + a_5 \sigma^3$	C
29	Boutora et al. (2022) [1]	$H/H_0 = a T^2 + b RH + c$	C
30	Banihashemi Dehkordi et al. (2022) [3]	Ångström–Prescott with metaheuristic calibration	A
31	Lan et al. (2023) [2]	sunshine models modified by the declination angle	C
32	Yu et al. (2023) [34]	zone-wise calibration of the Ångström–Prescott coefficients	A
33	Benmouiza (2023) [4]	kriging of sunshine duration over 56 Algerian stations	A
34	Djeldjli et al. (2024) [35]	hybrid firefly-optimised artificial neural network	D
35	Oulimar et al. (2024) [36]	statistical characterisation of the Adrar radiation regime	A
36	Bouramdane et al. (2026) [37]	validation of empirical models on inclined planes, Ghardaïa	C
37	Hui et al. (2026) [38]	nationwide zoning; empirical versus deep-learning comparison	D
38	Present work	$H/H_0 = a \sigma + b T^2 + c RH + d$	C

Three observations follow from Table 3. First, the sunshine-based family remains the most populated, and the most recent contributions refine the way its coefficients are estimated – by metaheuristic optimisation [3], by climatic zoning [34] or by an astronomical modulation of the coefficients themselves [2] – rather than abandoning it. Second, the temperature-based family, which developed mainly between 2004 and 2018, was explicitly motivated by the unavailability of sunshine records; the model of [1] belongs to that logic. Third, the data-driven family now reaches accuracies that the empirical correlations cannot match – Hui et al. [38] report $R^2 = 0.939$ for a long short-term memory network at national scale, and Djeldjli et al. [35] a correlation coefficient of 0.932 over six southern Algerian cities – but at the cost of a large training set, a loss of physical interpretability and coefficients that cannot be transcribed in a design handbook. For a site whose record is measured in tens of monthly values, a parsimonious and interpretable correlation remains the appropriate tool.

4. THEORETICAL BACKGROUND

4.1 Extraterrestrial radiation and day length

All the correlations reviewed above are written in terms of the clearness index, that is, of the ratio between the irradiation received at the ground and the irradiation that would be received at the top of the atmosphere on the same horizontal surface. The latter is purely astronomical and is computed, following Duffie and Beckman [39], from the solar declination, the sunset hour angle and the eccentricity correction factor. The declination is obtained from Cooper's equation:

$$\delta = 23.45 \sin [360 (284 + n) / 365] \quad (1)$$

where n is the day of the year counted from the first of January. The sunset hour angle follows from the geometry of the sphere:

$$\omega_s = \cos^{-1} (- \tan \varphi \cdot \tan \delta) \quad (2)$$

in which φ is the latitude of the site. It should be stressed that φ is the latitude and not the longitude: the two are easily interchanged in a spreadsheet, and the consequence is severe, since the seasonal amplitude of the extraterrestrial irradiation is governed almost entirely by this parameter. At the latitude of Tébéssa the correct computation yields $H_0 = 16.60 \text{ MJ}\cdot\text{m}^{-2}\cdot\text{day}^{-1}$ in December and $41.58 \text{ MJ}\cdot\text{m}^{-2}\cdot\text{day}^{-1}$ in June, a ratio of 2.5; any calculation producing a quasi-constant H_0 across the seasons should be regarded as erroneous. The astronomical day length is then

$$S_0 = (2/15) \omega_s \quad (3)$$

and the daily extraterrestrial irradiation on a horizontal surface is given by

$$H_0 = (24/\pi) I_{sc} [\cos \varphi \cos \delta \sin \omega_s + (\pi/180) \omega_s \sin \varphi \sin \delta] [1 + 0.033 \cos (360 n / 365)] \quad (4)$$

where $I_{sc} = 1367 \text{ W}\cdot\text{m}^{-2}$ is the solar constant. The monthly mean value of H_0 is evaluated for the representative day of each month proposed by Klein [40].

4.2 Clearness index and relative sunshine duration

The two dimensionless quantities on which the proposed model is built are the clearness index

$$K_T = H / H_0 \quad (5)$$

and the relative sunshine duration

$$\sigma = S / S_0 \quad (6)$$

in which S is the monthly mean daily sunshine duration in hours. Both quantities are bounded by zero and unity, which makes the regression coefficients comparable from one site to another and provides an immediate physical check on the data. Eq. (3), although written in [1], was not used there, the model being independent of the sunshine record; in the present formulation it supplies the denominator of the new predictor.

5. PROPOSED MODEL AND CALIBRATION STRATEGY

5.1 Formulation

The correlation proposed in [1] expresses the clearness index as a quadratic function of the air temperature and a linear function of the relative humidity. Its two predictors describe the thermal and hygrometric state of the boundary layer, but neither carries information on the duration of the direct beam. The natural extension is therefore to add the relative sunshine duration as a third predictor:

$$H / H_0 = a \sigma + b T^2 + c RH + d \quad (7)$$

where T is the monthly mean air temperature in degrees Celsius, RH the monthly mean relative humidity as a dimensionless fraction, and σ the relative sunshine duration of Eq. (6). The quadratic form in T is retained from [1] for continuity. Eq. (7) contains four unknown coefficients instead of three, which is not merely arithmetical: as shown in Section 5.2, it allows the whole calendar year to be used in the seasonal calibration instead of only three quarters of it.

In order to establish whether any gain in accuracy is attributable to the new predictor rather than to the increase in the number of degrees of freedom, five competing correlations are calibrated on the same data and assessed with the same indicators. They are listed in Table 4.

Table 4
Candidate models calibrated and compared in this study

Model	Analytical form	Predictors	Coefficients
M1	$H/H_0 = a T^2 + b RH + c$	T, RH	3
M2	$H/H_0 = a + b \sigma$	σ	2
M3 (proposed)	$H/H_0 = a \sigma + b T^2 + c RH + d$	σ, T, RH	4
M4	$H/H_0 = a \sigma + b \sigma^2 + c T^2 + d RH + e$	σ, T, RH	5
M5	$H/H_0 = a \sigma + b T^2 + c RH + d (\sigma \cdot RH) + e$	σ, T, RH	5
M6	$H/H_0 = a \sigma + b \sqrt{\Delta T} + c RH + d$	$\sigma, \Delta T, RH$	4

M1 reproduces the previous model [1] and serves as the reference against which the improvement is measured. M2 is the classical Ångström–Prescott correlation and provides the lower bound of complexity. M4 tests whether a quadratic dependence on σ , of the kind proposed by Ögelman et al. [12] and Aksoy [18], brings any additional benefit. M5 introduces an interaction term between sunshine and humidity, motivated by the fact that a humid atmosphere attenuates the beam more strongly for a given sunshine duration. M6 substitutes the square root of the diurnal temperature range for the quadratic temperature term, following Hassan et al. [29], and is intended as a fallback for sites where the sunshine record is short but the thermal record is complete.

5.2 Calibration by the method of determinants

The procedure used in [1] consisted in writing one equation per season, the number of seasons being equal to the number of unknowns, and in solving the resulting linear system by Cramer's rule. This procedure is retained here for continuity, and the passage from three to four unknowns has a welcome side effect: with three coefficients only three quarters of the year could be used and the choice of the omitted quarter was arbitrary, whereas with four coefficients the four quarters — December to February, March to May, June to August and September to November — all enter the system, so that the calibration rests on the complete annual cycle. Denoting by σ_i , T_i , RH_i and K_{Ti} the seasonal means of the i -th quarter, the system reads

$$a \sigma_i + b T_i^2 + c RH_i + d = K_{Ti}, i = 1 \dots 4 \quad (8)$$

and its solution is obtained from the ratios of the determinants of the four matrices formed by successively replacing each column of the coefficient matrix by the right-hand side vector. For the record analysed here the determinant of the system is 1.467, so the system is well conditioned in the algebraic sense.

5.3 Calibration by ordinary least squares and cross-validation

The determinant method has an intrinsic limitation which must be acknowledged: it produces an exact fit through as many points as there are unknowns, so that no residual exists, no confidence interval can be attached to the coefficients and no goodness-of-fit statistic can be computed. It is the solution of a linear system, not a regression. The coefficients of Eq. (7) are therefore also estimated by ordinary least squares over the twelve monthly values, the estimator being

$$\beta = (X^T X)^{-1} X^T y \quad (9)$$

in which X is the design matrix and y the vector of observed clearness indices. This route delivers the standard error of each coefficient, its Student statistic and the associated p-value, and it makes it possible to verify by the variance inflation factor whether the predictors can be estimated independently. Because the record is a climatological normal (Section 2.3), the predictive performance is assessed by leave-one-month-out cross-validation: for each month in turn the correlation is refitted on the eleven remaining months and used to predict the excluded one, so that every value reported in Section 6.4 is an out-of-sample estimate.

5.4 Statistical indicators

The single relative deviation used in [1] conflates the systematic bias and the random scatter and does not allow comparison with the international literature. The assessment carried out here relies on the ten indicators of Table 5. The root mean square error and its normalised form quantify the overall error; the mean bias error and the mean percentage error isolate the systematic component; the coefficient of

determination and the Nash–Sutcliffe efficiency [41] measure the fraction of variance explained; the t-statistic of Stone [42] indicates whether the bias is statistically significant; the expanded uncertainty gives the 95% confidence band; and the global performance indicator [43] aggregates the previous indicators into a single ranking criterion.

Table 5
Statistical indicators used for the assessment of the models

Indicator	Definition	Ideal value
RMSE	$[\sum (H_c - H_m)^2 / N]^{0.5}$	0
nRMSE	$100 \cdot \text{RMSE} / \text{mean}(H_m)$	0 (< 10% excellent)
MBE	$\sum (H_c - H_m) / N$	0
MAPE	$(100/N) \sum H_c - H_m / H_m$	0
MPE	$(100/N) \sum (H_c - H_m) / H_m$	0
R ²	coefficient of determination	1
NSE	$1 - \sum (H_c - H_m)^2 / \sum (H_m - \text{mean } H_m)^2$	1
t-stat	$[(N - 1) \text{MBE}^2 / (\text{RMSE}^2 - \text{MBE}^2)]^{0.5}$	< 2.20 (t critical, 11 d.o.f., 95%)
U95	$1.96 (\text{SD}^2 + \text{RMSE}^2)^{0.5}$	0
GPI	aggregated normalised score	maximum

H_m denotes the value taken from the CDER database, H_c the value computed by the correlation, N = 12 the number of monthly values and SD the standard deviation of the residuals.

6. RESULTS AND ANALYSIS

6.1 Climatic characterisation of the site

Table 6 gives the monthly means of the four measured variables together with the astronomical quantities computed from Eqs. (1) to (4), and Fig. 1 displays them. The extraterrestrial irradiation rises from 16.60 MJ·m⁻²·day⁻¹ in December to 41.58 MJ·m⁻²·day⁻¹ in June and the day length varies between 9.65 and 14.35 hours; these two columns depend only on the latitude and can be used to verify the computation chain before any meteorological data are introduced.

Table 6
Monthly means of the astronomical and meteorological variables at Tébessa, 2015–2024

Month	n	H0	S0 (h)	T (°C)	ΔT (°C)	RH (–)	S (h)	σ (–)	H	KT (–)
January	17	18.02	9.90	7.20	10.47	0.679	6.70	0.677	10.66	0.591
February	47	23.11	10.75	8.14	10.76	0.649	7.36	0.685	13.76	0.595
March	75	29.45	11.77	11.14	11.46	0.582	8.04	0.683	17.63	0.598
April	105	35.76	12.90	14.35	12.38	0.523	9.04	0.701	21.40	0.599
May	135	39.96	13.87	19.15	14.08	0.462	10.23	0.738	24.16	0.604
June	162	41.58	14.35	24.14	15.18	0.360	11.66	0.812	26.46	0.636
July	198	40.69	14.13	27.32	15.85	0.308	12.05	0.853	27.01	0.664
August	228	37.30	13.31	26.34	14.79	0.348	11.27	0.847	24.66	0.661
September	258	31.67	12.21	22.45	12.88	0.501	9.14	0.749	20.17	0.637
October	288	24.93	11.08	17.95	11.69	0.568	7.79	0.703	15.66	0.628
November	318	19.19	10.12	12.45	10.67	0.640	6.80	0.672	11.53	0.601
December	344	16.60	9.65	8.46	9.59	0.701	6.36	0.659	9.92	0.598
Year	—	29.85	12.04	16.59	12.48	0.527	8.87	0.732	18.58	0.618

H0 and H are expressed in MJ·m⁻²·day⁻¹. The last line gives the annual means; H0 is there the mean of the twelve monthly values.

6.2 Correlation structure of the predictors

Table 7 reports the matrix of linear correlation coefficients and the variance inflation factors, and Fig. 2 displays the matrix. The first expected result is confirmed: the relative sunshine duration is strongly correlated with the clearness index, $r = +0.921$, which is the physical premise of the proposed model. But the second result is more consequential. The predictors are not merely correlated with the target; they are almost perfectly correlated with one another. The coefficient between σ and RH is -0.962 , that between T and σ is $+0.930$, and the variance inflation factors reach 181 for σ , 634 for T^2 and 142 for RH, against the conventional warning threshold of ten.

The reason is physical rather than accidental. Over a semi-arid plateau every one of these variables is driven by the same annual solar cycle: the days lengthen, the sky clears, the air warms and dries, all in phase. On a monthly climatological normal, from which the inter-annual and day-to-day fluctuations have been averaged out, nothing remains that could separate them. This observation governs the interpretation of everything that follows.

Table 7
Correlation matrix of the predictors and variance inflation factors

σ	σ	T	ΔT	RH	KT	VIF
	1.000	+ 0.930	+ 0.946	-0.962	+ 0.921	180.9
T		1.000	+ 0.930	-0.949	+ 0.930	379.7
T ²	+ 0.966	+ 0.994	+ 0.933	-0.954	+ 0.958	633.5
ΔT			1.000	-0.991	+ 0.812	61.2
RH				1.000	-0.856	141.6
KT					1.000	—

Figure 3 shows the clearness index against the relative sunshine duration and Fig. 4 compares the three predictors on the same vertical scale. The simple regression of KT on σ gives $R^2 = 0.849$, but T² alone gives $R^2 = 0.918$ and RH alone $R^2 = 0.733$. On this data set the quadratic temperature term is therefore, taken in isolation, a marginally better single predictor of the clearness index than the sunshine ratio — a result that must be stated plainly, and that Section 7 interprets.

6.3 Calibrated coefficients

Table 8 gives the coefficients of the three principal models estimated by ordinary least squares, with their standard errors, Student statistics and p-values.

Table 8
Coefficients of models M1, M2 and M3 estimated by ordinary least squares

Model	Coefficient	Estimate	Standard error	t	p-value
M1	b (T ²)	0.000160	0.000025	6.48	< 0.001
	c (RH)	0.119249	0.046529	2.56	0.010
	constant	0.502952	0.032288	15.58	< 0.001
M2	b (σ)	0.349923	0.046649	7.50	< 0.001
	constant	0.361765	0.034264	10.56	< 0.001
M3	a (σ)	0.149002	0.125262	1.19	0.234
	b (T ²)	0.000137	0.000031	4.44	< 0.001
	c (RH)	0.152225	0.053273	2.86	0.004
	d (constant)	0.384017	0.104851	3.66	< 0.001

The proposed correlation for Tébessa therefore takes the final form

$$H / H_0 = 0.14900 \sigma + 0.000137 T^2 + 0.15222 RH + 0.38402 \quad (10)$$

Two features of Table 8 require comment. The first is that the coefficient of σ in M3, although positive and of the expected sign, carries a standard error of the same order as itself and is not significant at the 5% level ($p = 0.234$), whereas the same coefficient in M2, where σ is the only predictor, is highly significant ($p < 0.001$). This is the classical signature of multicollinearity: the information carried by σ is largely already contained in T^2 and RH, so that the regression cannot attribute it unambiguously. The second is that the coefficient of RH is positive in both M1 and M3, which is physically counter-intuitive — a more humid atmosphere should attenuate the beam, not enhance it. It should be read as a partial coefficient conditional on the other terms and not as a physical response; the correlation between RH and KT taken alone is, as expected, negative (-0.856).

Table 9 compares the least-squares estimates with those obtained by the determinant method applied to the four seasonal means.

Table 9
Coefficients of the proposed model: least squares against the determinant method

Coefficient	Ordinary least squares (12 monthly values)	Cramer 4 × 4 (4 seasonal means)	Relative difference
a (σ)	0.149002	0.032581	−78%
b (T^2)	0.000137	0.000166	+ 21%
c (RH)	0.152225	0.140401	−8%
d (constant)	0.384017	0.467400	+ 22%

The agreement is good for the humidity and temperature terms but poor for the sunshine term, whose determinant estimate is smaller by nearly four fifths. This is the expected consequence of an exact fit through four averaged points: the seasonal averaging removes precisely the variability on which the least-squares estimate of the weakest coefficient depends. The least-squares values are therefore retained as the reference, and the determinant solution is reported for continuity with [1] and as a sensitivity check.

6.4 Predictive performance

Table 10 gathers the ten indicators computed by leave-one-month-out cross-validation, so that every value is an out-of-sample estimate. Figure 5 compares the measured monthly irradiation with that estimated by the previous and the proposed correlations, Fig. 6 gives the one-to-one plots of the six candidates, and Fig. 8 summarises three indicators graphically.

Table 10
Statistical performance of the candidate models, leave-one-month-out cross-validation

Model	RMSE	nRMSE (%)	MBE	MAPE (%)	MPE (%)	R ²	t-stat	U95	GPI
M1 — previous model [1]	0.2537	1.365	+ 0.0072	0.948	+ 0.006	0.9982	0.094	0.719	-0.164
M2 — Ångström–Prescott	0.3546	1.908	+ 0.0245	1.545	+ 0.008	0.9965	0.230	1.004	-4.635
M3 — proposed	0.2430	1.308	+ 0.0032	1.018	+ 0.004	0.9984	0.043	0.689	+ 0.215
M4	0.2073	1.116	+ 0.0040	0.960	+ 0.025	0.9988	0.065	0.588	+ 1.189
M5	0.2383	1.283	+ 0.0050	1.116	+ 0.039	0.9984	0.069	0.675	+ 0.041
M6	0.3231	1.738	+ 0.0589	1.452	+ 0.283	0.9971	0.616	0.908	-5.537
M3 by Cramer 4 × 4	0.1935	1.041	+ 0.0466	0.721	+ 0.200	0.9990	0.824	0.540	-0.484

RMSE, MBE and U95 are expressed in MJ·m⁻²·day⁻¹. The Cramer solution is not cross-validated — it is an exact fit through four seasonal means and its indicators are therefore in-sample; it is listed for comparison only and must not be read as an out-of-sample performance.

The proposed correlation improves on the previous one in every respect, but by margins that must be stated precisely. The root mean square error falls from 0.2537 to 0.2430 MJ·m⁻²·day⁻¹, a reduction of 4.2%; the expanded uncertainty falls by the same proportion; the systematic bias falls from + 0.0072 to + 0.0032 MJ·m⁻²·day⁻¹, a reduction of 56%; and the Student statistic of Stone falls from 0.094 to 0.043, both being far below the critical value of 2.20 at eleven degrees of freedom. Against the classical Ångström–Prescott correlation the gain is much larger: 31.5% in root mean square error. The mean absolute percentage error is the one indicator that does not improve, rising slightly from 0.948% to 1.018%, which reflects a redistribution of the error between months rather than an overall degradation, as Section 6.5 shows.

Models M4 and M5, which add one coefficient each, reach a lower root mean square error than M3 — 0.2073 and 0.2383 MJ·m⁻²·day⁻¹ respectively — and M4 obtains the highest global performance indicator of the set. It is nevertheless not adopted here, for a reason that Table 8 makes clear by analogy: the coefficients of M4 are $a = -1.436$ for σ and $+0.961$ for σ^2 , that is, a large negative linear term compensated by a large positive quadratic one over a range of σ spanning barely 0.19. Such a pair is a numerical artefact of fitting five coefficients to twelve collinear points; it has no physical reading, and a

correlation whose coefficients change sign under a small perturbation of the calibration set cannot be recommended for use at other sites. M3 is therefore retained on grounds of parsimony and stability, and M4 is reported for completeness.

6.5 Seasonal behaviour of the residuals

A correlation may be satisfactory on the annual average and yet fail systematically in a particular season, which for a solar application is a more serious defect than a slightly larger overall scatter. Table 11 reports the bias and the mean absolute percentage error season by season, and Fig. 7 shows the monthly residuals of the two correlations side by side.

Table 11
Seasonal decomposition of the error

Season	Months	MBE – M1	MAPE – M1 (%)	MBE – M3	MAPE – M3 (%)
Winter	December – February	-0.0257	0.315	+ 0.0202	0.363
Spring	March – May	+ 0.1007	1.038	+ 0.0398	1.075
Summer	June – August	-0.0888	0.655	-0.0285	0.486
Autumn	September – November	+ 0.0429	0.855	-0.0143	0.808

The pattern is instructive, and it is not the one that the working hypothesis of Section 1 would have predicted. The introduction of the sunshine term reduces the absolute bias in three seasons out of four – by 60% in spring, 68% in summer and 67% in autumn – and reduces the mean absolute percentage error in summer, from 0.655% to 0.486%, and in autumn. In winter, however, the bias changes sign and the percentage error increases slightly. The improvement is therefore concentrated in the warm and clear half of the year rather than in the cloudy half, which is the opposite of what a purely cloud-related interpretation would suggest. A plausible reading is that in winter, when the clearness index is at its flattest – it varies by only 0.008 between December and February – there is almost no signal for any predictor to capture, so that the additional term can only add estimation noise.

6.6 Sensitivity analysis

Each predictor was perturbed in turn by $\pm 10\%$ around its annual mean, the other two being held at their mean values, and the resulting variation of the clearness index was recorded. The results are given in Table 12 and Fig. 9.

Table 12
Sensitivity of the clearness index to a $\pm 10\%$ perturbation of each predictor

Predictor	Annual mean	ΔKT for -10% (%)	ΔKT for $+10\%$ (%)	Rank
σ	0.732	-1.78	+ 1.78	1
RH	0.527	-1.31	+ 1.31	2
T	16.59°C	-1.18	+ 1.30	3

The relative sunshine duration is the most influential of the three predictors of Eq. (10): a 10% change in σ moves the clearness index by 1.78%, against 1.31% for the relative humidity and about 1.24% on average for the temperature. The response to σ and to RH is symmetrical because both enter linearly, whereas the response to T is slightly asymmetrical because it enters quadratically. This result is worth setting beside the significance test of Table 8: σ carries the largest physical leverage in the fitted equation, yet its coefficient is the least precisely determined. The two statements are not contradictory – the first concerns the magnitude of the effect, the second the precision with which the data allow it to be measured – and together they summarise the position of this study.

7. DISCUSSION

Four points deserve discussion.

The first concerns what the results do and do not establish. Adding the relative sunshine duration to the temperature–humidity correlation of [1] improves the estimate: the bias is more than halved, the expanded uncertainty and the root mean square error fall by about four per cent, and every indicator except the mean absolute percentage error moves in the right direction. The term should therefore be retained. But the improvement in root mean square error is modest, and the coefficient of σ is not individually significant on this data set. It would be easy, and wrong, to present the 56% reduction in bias without the 4.2% reduction in scatter, or to quote the coefficient without its p-value. The honest summary is that the sunshine term helps, that it helps mainly by removing systematic bias, and that the present data cannot demonstrate more than that.

The second concerns the reason, which is quantified in Section 6.2 rather than assumed. On a monthly climatological normal at a semi-arid site the three predictors are almost perfectly collinear: variance inflation factors of 181, 634 and 142 against a warning threshold of ten, and a correlation of -0.962 between σ and RH. All three variables are driven by the same annual solar cycle, and once the inter-annual and day-to-day fluctuations have been averaged away nothing remains to distinguish them. In such a configuration a regression can reproduce the observations very accurately – R^2 exceeds 0.998 for five of the six candidates – while being unable to apportion the credit among the predictors. The apparently paradoxical finding of Fig. 4, that T^2 alone explains the clearness index slightly better than σ alone, should be read in this light: it does not mean that temperature causes transmissivity, only that at monthly-normal resolution the two are indistinguishable proxies for the season.

The third concerns the practical consequence, which is a recommendation about data rather than about models. A decisive test of the sunshine term requires calibration data in which σ and T can vary independently, and that means daily values, or at the very least individual-year monthly values rather than a ten-year normal. On daily data a cloudy day in July and a clear day in July have very different values of σ and nearly the same value of T , and the regression can then separate them; this is precisely the configuration in which the sunshine-based correlations reviewed in Section 3 obtain their advantage. The satellite product SARA3 [5], which delivers daily sunshine duration at 0.05° resolution over the whole of Algeria, makes such a data set available for Tébessa at no cost, and its use is the natural continuation of this work.

The fourth concerns transferability. The correlation proposed here is calibrated for a BSk climate at 850 m, and its coefficients absorb the local optical properties of the atmosphere. The assertion made in [1], that a model calibrated at one Algerian site may be used anywhere in the world where the climate is similar, should be qualified: what the present results support is the bounded statement that Eq. (10) is applicable to semi-arid continental sites of the Maghreb lying between 30° and 37° N at elevations of 600 to 1200 m, and that it should be re-calibrated elsewhere. A regional extension by kriging of the coefficients, along the lines of the sunshine mapping carried out by Benmouiza [4] over 56 Algerian stations, would be the appropriate way to generalise it.

A word must finally be said about the alternative offered by machine learning. The recent literature leaves no doubt that neural architectures outperform empirical correlations when a sufficiently long training record is available: Djeldjli et al. [35] obtain a correlation coefficient of 0.932 over six southern Algerian cities with a firefly-optimised network, and Hui et al. [38] report $R^2 = 0.939$ for a long short-term memory model at national scale. Those approaches require data volumes that a twelve-point climatological normal cannot supply, they do not yield transferable coefficients, and they cannot be inspected physically. The two families answer different questions, and the correlation proposed here is intended for the situation — still the most common one in Algeria — in which a designer possesses a monthly climatic normal and nothing more.

8. CONCLUSION

The estimation of the global solar radiation from ordinary meteorological variables remains, in a country whose radiometric network covers only a small number of sites, the practical route to the assessment of the solar resource. The present work took as its starting point a documented weakness of a previously published correlation — the degradation of its accuracy over the continental and elevated sites of eastern Algeria — and tested the hypothesis that the weakness stems from the absence of any direct descriptor of atmospheric transmissivity among its predictors.

A four-parameter correlation was accordingly proposed, in which the relative sunshine duration is added to the quadratic temperature and linear humidity terms of the earlier formulation. It was calibrated for the wilaya of Tébessa on the CDER radiation database completed by the ONM meteorological record,

assessed by leave-one-month-out cross-validation and compared with five competing correlations by means of ten statistical indicators. The main outcomes are the following:

1. the calibrated correlation for Tébessa is $H/H_0 = 0.14900 \sigma + 0.000137 T^2 + 0.15222 RH + 0.38402$, with $R^2 = 0.9984$, $RMSE = 0.243 \text{ MJ}\cdot\text{m}^{-2}\cdot\text{day}^{-1}$ and $nRMSE = 1.31\%$ in cross-validation;
2. with respect to the temperature–humidity model of [1] the systematic bias is reduced by 56% (from $+0.0072$ to $+0.0032 \text{ MJ}\cdot\text{m}^{-2}\cdot\text{day}^{-1}$) and the Student statistic from 0.094 to 0.043, while the root mean square error and the expanded uncertainty each improve by 4.2%; with respect to the classical Ångström–Prescott correlation the root mean square error improves by 31.5%;
3. the improvement is concentrated in the warm half of the year, the absolute bias falling by 60 to 68% in spring, summer and autumn, whereas in winter, where the clearness index varies by only 0.008 between December and February, the additional term brings no benefit;
4. the sensitivity analysis ranks the relative sunshine duration first among the three predictors, a $\pm 10\%$ perturbation moving the clearness index by $\pm 1.78\%$ against $\pm 1.31\%$ for the relative humidity and $\pm 1.24\%$ for the temperature;
5. the coefficient of σ is nevertheless not significant at the 5% level ($p = 0.234$), because on a monthly climatological normal the three predictors are almost perfectly collinear – variance inflation factors of 181, 634 and 142, and a correlation of -0.962 between σ and RH.

The recommendation that follows is twofold. For immediate engineering use at Tébessa and at comparable semi-arid highland sites, Eq. (10) may be adopted: it is more accurate and markedly less biased than its predecessor, and its four coefficients are stable. For the scientific question of how much the sunshine term is really worth, the present data are not conclusive, and the answer requires a calibration set in which sunshine and temperature vary independently – that is, daily data. The satellite sunshine duration product of CM SAF SARA-H3 provides such a data set for the whole of Algeria at no cost, and its exploitation, together with a regional extension of the coefficients by kriging over the eastern high plains, constitutes the natural continuation of this work.

ABBREVIATIONS

H	monthly mean daily global solar irradiation on a horizontal surface, $\text{MJ}\cdot\text{m}^{-2}\cdot\text{day}^{-1}$
H0	monthly mean daily extraterrestrial irradiation on a horizontal surface, $\text{MJ}\cdot\text{m}^{-2}\cdot\text{day}^{-1}$
Hc	irradiation computed by the correlation, $\text{MJ}\cdot\text{m}^{-2}\cdot\text{day}^{-1}$
Hm	irradiation taken from the CDER database, $\text{MJ}\cdot\text{m}^{-2}\cdot\text{day}^{-1}$
Isc	solar constant, $1367 \text{ W}\cdot\text{m}^{-2}$
KT	clearness index, $H/H0$, dimensionless
N	number of monthly values, $N = 12$
n	day of the year counted from 1 January
RH	monthly mean relative humidity, dimensionless fraction
S	monthly mean daily sunshine duration, h
S0	astronomical day length, h
T	monthly mean air temperature, $^{\circ}\text{C}$
Tmax, Tmin	monthly mean daily maximum and minimum air temperature, $^{\circ}\text{C}$
X, y	design matrix and response vector of the regression
Z	elevation above sea level, m
Greek symbols	
β	vector of regression coefficients
δ	solar declination, $^{\circ}$
ΔT	diurnal temperature range, $T_{\text{max}} - T_{\text{min}}$, $^{\circ}\text{C}$
σ	relative sunshine duration, $S/S0$, dimensionless
φ	latitude of the site, $^{\circ}$
ω_s	sunset hour angle, $^{\circ}$
Abbreviations	
CDER	Centre de Développement des Energies Renouvelables
CM SAF	Satellite Application Facility on Climate Monitoring
GPI	global performance indicator
GSR	global solar radiation
MBE	mean bias error

MAPE	mean absolute percentage error
MPE	mean percentage error
NSE	Nash–Sutcliffe efficiency
ONM	Office National de la Météorologie
RMSE	root mean square error
SARAH	Surface Solar Radiation Data Set – Heliosat
SDU	sunshine duration product of CM SAF
VIF	variance inflation factor

DECLARATIONS

Funding

No funding was received for conducting this study

Author Contribution

Boutora Tidjani was responsible for the conceptualization of the hybrid model, data collection from the Renewable Energy Development Centre and the Algerian National Meteorological Office, statistical analysis (including multicollinearity and variance inflation factor evaluations).

ACKNOWLEDGMENT

The authors thank the Renewable Energy Development Centre (CDER) and the Algerian National Meteorological Office (ONM) for providing the data used in this study, and EUMETSAT for the free distribution of the CM SAF SARAH-3 climate data record.

Data Availability

All datasets generated or analyzed during this study are included in this published article.

REFERENCES

1. Boutora, T., Rebah, N. & Dib, D. A new model to predict the global solar radiation GSR of Souk-Ahras city. *Eur. J. Electr. Eng.* **24** (2), 67–76. <https://doi.org/10.18280/ejee.240201> (2022).

2. Lan, K. et al. The applicability of sunshine-based global solar radiation models modified with meteorological factors for different climate zones of China. *Front. Energy Res.* **10**, 1010745. <https://doi.org/10.3389/fenrg.2022.1010745> (2023).
3. Banihashemi Dehkordi, S. N., Bakhtiari, B., Qaderi, K. & Ahmadi, M. M. Calibration and validation of the Angstrom-Prescott model in solar radiation estimation using optimization algorithms. *Sci. Rep.* **12**, 4855. <https://doi.org/10.1038/s41598-022-08744-6> (2022).
4. Benmouiza, K. Solar zoning maps of Algeria based on sunshine duration data and kriging method. *Int. J. Heat Technol.* **41** (3). <https://doi.org/10.18280/ijht.410317> (2023).
5. Pfeifroth, U. et al. SARA-3 - satellite-based climate data records of surface solar radiation. *Earth Syst. Sci. Data.* **16**, 5243–5265. <https://doi.org/10.5194/essd-16-5243-2024> (2024).
6. Ångström, A. Solar and terrestrial radiation. *Q. J. R. Meteorol. Soc.* **50** (210), 121–126. <https://doi.org/10.1002/qj.49705021008> (1924).
7. Glover, J. & McCulloch, J. S. G. The empirical relation between solar radiation and hours of sunshine. *Q. J. R. Meteorol. Soc.* **84** (360), 172–175. <https://doi.org/10.1002/qj.49708436011> (1958).
8. Page, J. K. The estimation of monthly mean values of daily total short wave radiation on vertical and inclined surfaces from sunshine records for latitudes 40°N-40°S. In Proceedings of the United Nations Conference on New Sources of Energy, 98(4): 378–390. (1961).
9. Swartman, R. K. & Ogunlade, O. Solar radiation estimates from common parameters. *Sol. Energy.* **11** (3–4), 170–172. [https://doi.org/10.1016/0038-092X\(67\)90026-6](https://doi.org/10.1016/0038-092X(67)90026-6) (1967).
10. Iqbal, M. Correlation of average diffuse and beam radiation with hours of bright sunshine. *Sol. Energy.* **23** (2), 169–173. [https://doi.org/10.1016/0038-092X\(79\)90118-X](https://doi.org/10.1016/0038-092X(79)90118-X) (1979).
11. Khogali, A., Ramadan, M. R. I., Ali, Z. E. H. & Fattah, Y. A. Global and diffuse solar irradiance in Yemen (YAR). *Sol. Energy.* **31** (1), 55–62. [https://doi.org/10.1016/0038-092X\(83\)90033-6](https://doi.org/10.1016/0038-092X(83)90033-6) (1983).
12. Ögelman, H., Ecevit, A. & Tasdemiroğlu, E. A new method for estimating solar radiation from bright sunshine data. *Sol. Energy.* **33** (6), 619–625. [https://doi.org/10.1016/0038-092X\(84\)90018-5](https://doi.org/10.1016/0038-092X(84)90018-5) (1984).
13. Bahel, V., Bakhsh, H. & Srinivasan, R. A correlation for estimation of global solar radiation. *Energy* **12** (2), 131–135. [https://doi.org/10.1016/0360-5442\(87\)90117-4](https://doi.org/10.1016/0360-5442(87)90117-4) (1987).
14. Gopinathan, K. K. A general formula for computing the coefficients of the correlation connecting global solar radiation to sunshine duration. *Sol. Energy.* **41** (6), 499–502. [https://doi.org/10.1016/0038-092X\(88\)90052-7](https://doi.org/10.1016/0038-092X(88)90052-7) (1988).
15. Newland, F. J. A study of solar radiation models for the coastal region of South China. *Sol. Energy.* **43** (4), 227–235. [https://doi.org/10.1016/0038-092X\(89\)90022-4](https://doi.org/10.1016/0038-092X(89)90022-4) (1989).
16. Akinoğlu, B. G. & Ecevit, A. A further comparison and discussion of sunshine-based models to estimate global solar radiation. *Energy* **15** (10), 865–872. [https://doi.org/10.1016/0360-5442\(90\)90068-D](https://doi.org/10.1016/0360-5442(90)90068-D) (1990).
17. Samuel, T. D. M. A. Estimation of global radiation for Sri Lanka. *Sol. Energy.* **47** (5), 333–337. [https://doi.org/10.1016/0038-092X\(91\)90026-S](https://doi.org/10.1016/0038-092X(91)90026-S) (1991).

18. Aksoy, B. Estimated monthly average global radiation for Turkey and its comparison with observations. *Renew. Energy*. **10** (4), 625–633. [https://doi.org/10.1016/S0960-1481\(96\)00035-3](https://doi.org/10.1016/S0960-1481(96)00035-3) (1997).
19. Ampratwum, D. B. & Dorvlo, A. S. Estimation of solar radiation from the number of sunshine hours. *Appl. Energy*. **63** (3), 161–167. [https://doi.org/10.1016/S0306-2619\(99\)00025-2](https://doi.org/10.1016/S0306-2619(99)00025-2) (1999).
20. Chegaar, M. & Chibani, A. Global solar radiation estimation in Algeria. *Energy. Conv. Manag.* **42** (8), 967–973. [https://doi.org/10.1016/S0196-8904\(00\)00105-9](https://doi.org/10.1016/S0196-8904(00)00105-9) (2001).
21. Chen, R., Ersi, K., Yang, J., Lu, S. & Zhao, W. Validation of five global radiation models with measured daily data in China. *Energy. Conv. Manag.* **45** (11–12), 1759–1769. <https://doi.org/10.1016/j.enconman.2003.09.019> (2004).
22. Bakirci, K. Estimation of solar radiation by using ASHRAE clear-sky model in Erzurum, Turkey. *Energy Sources, Part A*, 31(3): 208–216. (2009). <https://doi.org/10.1080/15567030701522534>
23. Olayinka, S. Estimation of global and diffuse solar radiations for selected cities in Nigeria. *Int. J. Energy Environ. Eng.* **2** (3), 13–33 (2011).
24. Adaramola, M. S. Estimating global solar radiation using common meteorological data in Akure. *Nigeria Renew. Energy*. **47**, 38–44. <https://doi.org/10.1016/j.renene.2012.04.005> (2012).
25. Adhikari, K. R., Bhattarai, B. K. & Gurung, S. Estimation of global solar radiation for four selected sites in Nepal using sunshine hours, temperature and relative humidity. *J. Power Energy Eng.* **1** (3), 1–9. <https://doi.org/10.4236/jpee.2013.13003> (2013).
26. Li, M. F., Tang, X. P., Wu, W. & Liu, H. B. General models for estimating daily global solar radiation for different solar radiation zones in mainland China. *Energy. Conv. Manag.* **70**, 139–148. <https://doi.org/10.1016/j.enconman.2013.03.004> (2013).
27. Mecibah, M. S., Boukelia, T. E., Tahtah, R. & Gairaa, K. Introducing the best model for estimation the monthly mean daily global solar radiation on a horizontal surface (case study: Algeria). *Renew. Sustain. Energy Rev.* **36**, 194–202 (2014).
28. Yaiche, M. R., Bouhanik, A., Bekkouche, S. M. A., Malek, A. & Benouaz, T. Revised solar maps of Algeria based on sunshine duration. *Energy. Conv. Manag.* **82**, 114–123 (2014).
29. Hassan, G. E., Youssef, M. E., Mohamed, Z. E., Ali, M. A. & Hanafy, A. A. New temperature-based models for predicting global solar radiation. *Appl. Energy*. **179**, 437–450. <https://doi.org/10.1016/j.apenergy.2016.07.006> (2016).
30. Jahani, B., Dinpashoh, Y. & Nafchi, A. R. Evaluation and development of empirical models for estimating daily solar radiation. *Renew. Sustain. Energy Rev.* **73**, 878–891. <https://doi.org/10.1016/j.rser.2017.01.124> (2017).
31. Fan, J. et al. Evaluation and development of temperature-based empirical models for estimating daily global solar radiation in humid regions. *Energy* **144**, 903–914. <https://doi.org/10.1016/j.energy.2017.12.091> (2018).
32. Ekici, C. & Teke, I. Developing a new solar radiation estimation model based on Buckingham theorem. *Results Phys.* **9**, 263–269. <https://doi.org/10.1016/j.rinp.2018.02.064> (2018).

33. Yıldırım, H. B., Teke, A. & Antonanzas-Torres, F. Evaluation of classical parametric models for estimating solar radiation in the Eastern Mediterranean region of Turkey. *Renew. Sustain. Energy Rev.* **82**, 2053–2065. <https://doi.org/10.1016/j.rser.2017.08.033> (2018).
34. Yu, X. et al. Calibration of the Ångström-Prescott model for accurately estimating solar radiation spatial distribution in areas with few global solar radiation stations: a case study of the China tropical zone. *Atmosphere* **14** (12), 1825. <https://doi.org/10.3390/atmos14121825> (2023).
35. Djeldjli, H., Benatiallah, D., Tanougast, C. & Benatiallah, A. Solar radiation forecasting based on ANN, SVM and a novel hybrid FFA-ANN model: a case study of six cities south of Algeria. *AIMS Energy*. **12** (1), 62–83. <https://doi.org/10.3934/energy.2024004> (2024).
36. Oulimar, I., Bouchouicha, K., Bailek, N. & Bellaoui, M. Statistical study of global solar radiation in the Algerian desert: a case study of Adrar town. *Theoret. Appl. Climatol.* **155** (4), 3493–3504. <https://doi.org/10.1007/s00704-024-04834-9> (2024).
37. Bouramdane, A., Louazene, M. L. & Benmir, A. Validation and comparative assessment of empirical models for solar energy prediction on inclined planes: a case study in Ghardaïa, Algeria. *Int. J. Energetica*. **11** (1), 40–47. <https://doi.org/10.47238/ijeca.v11i1.317> (2026).
38. Hui, B. et al. Nationwide solar radiation zoning and performance comparison of empirical and deep learning models. *Appl. Sci.* **16** (9), 4229. <https://doi.org/10.3390/app16094229> (2026).
39. Duffie, J. A. & Beckman, W. A. *Solar Engineering of Thermal Processes* 4th edn (Wiley, 2013).
40. Klein, S. A. Calculation of monthly average insolation on tilted surfaces. *Sol. Energy*. **19** (4), 325–329 (1977).
41. Nash, J. E. & Sutcliffe, J. V. River flow forecasting through conceptual models part I - a discussion of principles. *J. Hydrol.* **10** (3), 282–290 (1970).
42. Stone, R. J. Improved statistical procedure for the evaluation of solar radiation estimation models. *Sol. Energy*. **51** (4), 289–291 (1993).
43. Despotovic, M., Nedic, V., Despotovic, D. & Cvetanovic, S. Review and statistical analysis of different global solar radiation sunshine models. *Renew. Sustain. Energy Rev.* **52**, 1869–1880 (2015).
44. World Meteorological Organization. *Guide to Instruments and Methods of Observation* (WMO-No. 8) (Volume I: Measurement of Meteorological Variables. WMO, 2021).
45. Renewable Energy Development. Centre (CDER). <http://www.cder.dz>
46. Algerian National Meteorological Office (ONM). <https://www.meteo.dz>
47. Note on the reference list. Entries [1]–[5], [34]–[38] were verified against the publishers' pages during the preparation of this manuscript. Entries [7]–[33] are carried over from the reference list of [1] and were therefore already checked at the time of its publication. Entries [6], [27], [28] and [40]–[44] are standard works whose bibliographic details should nevertheless be re-verified in Crossref before submission, since their DOIs are not reproduced here in full.

Figures

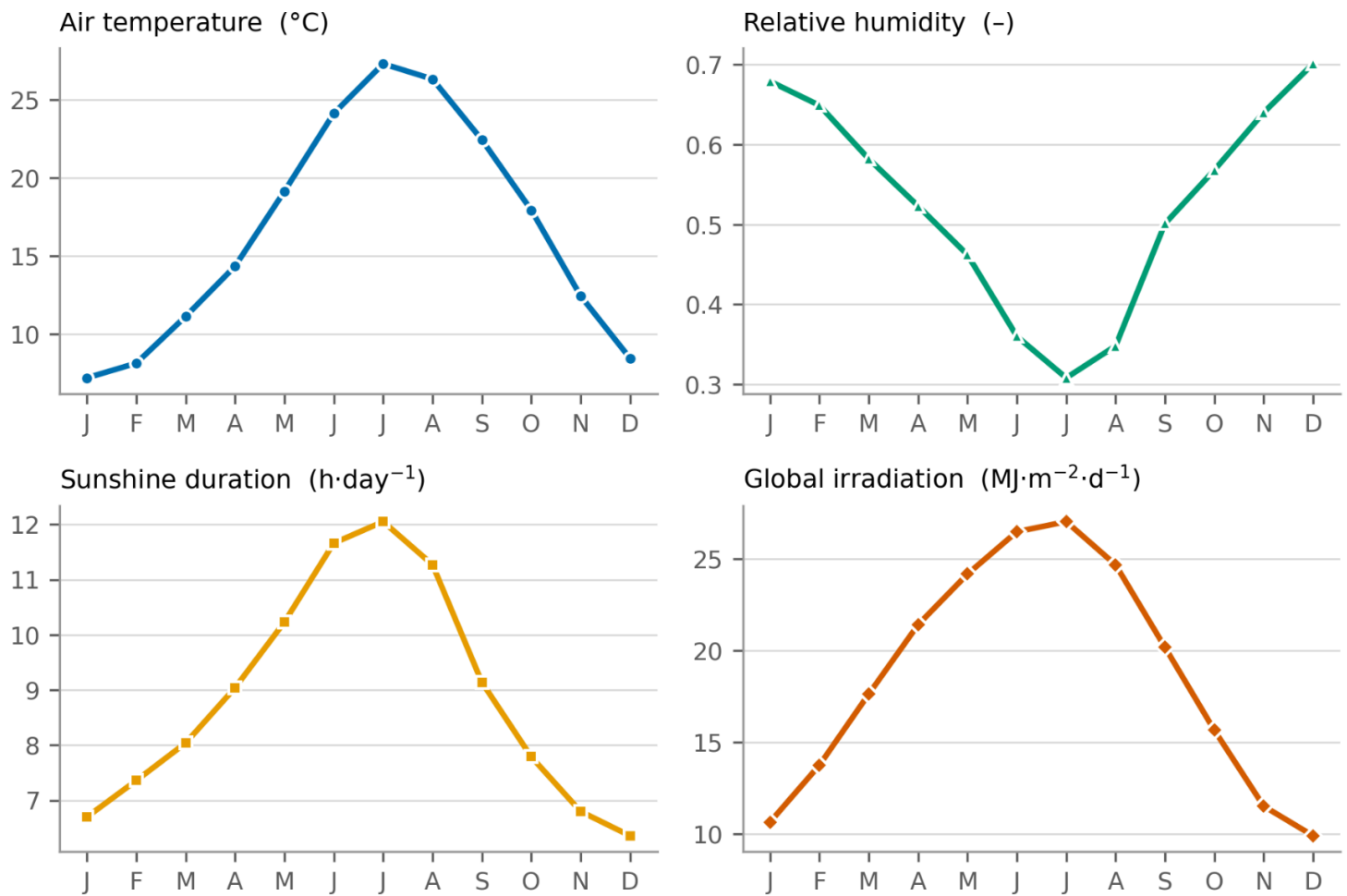


Figure 1

Monthly means of air temperature, relative humidity, sunshine duration and global irradiation at Tébessa, 2015–2024.

Correlation matrix of the predictors

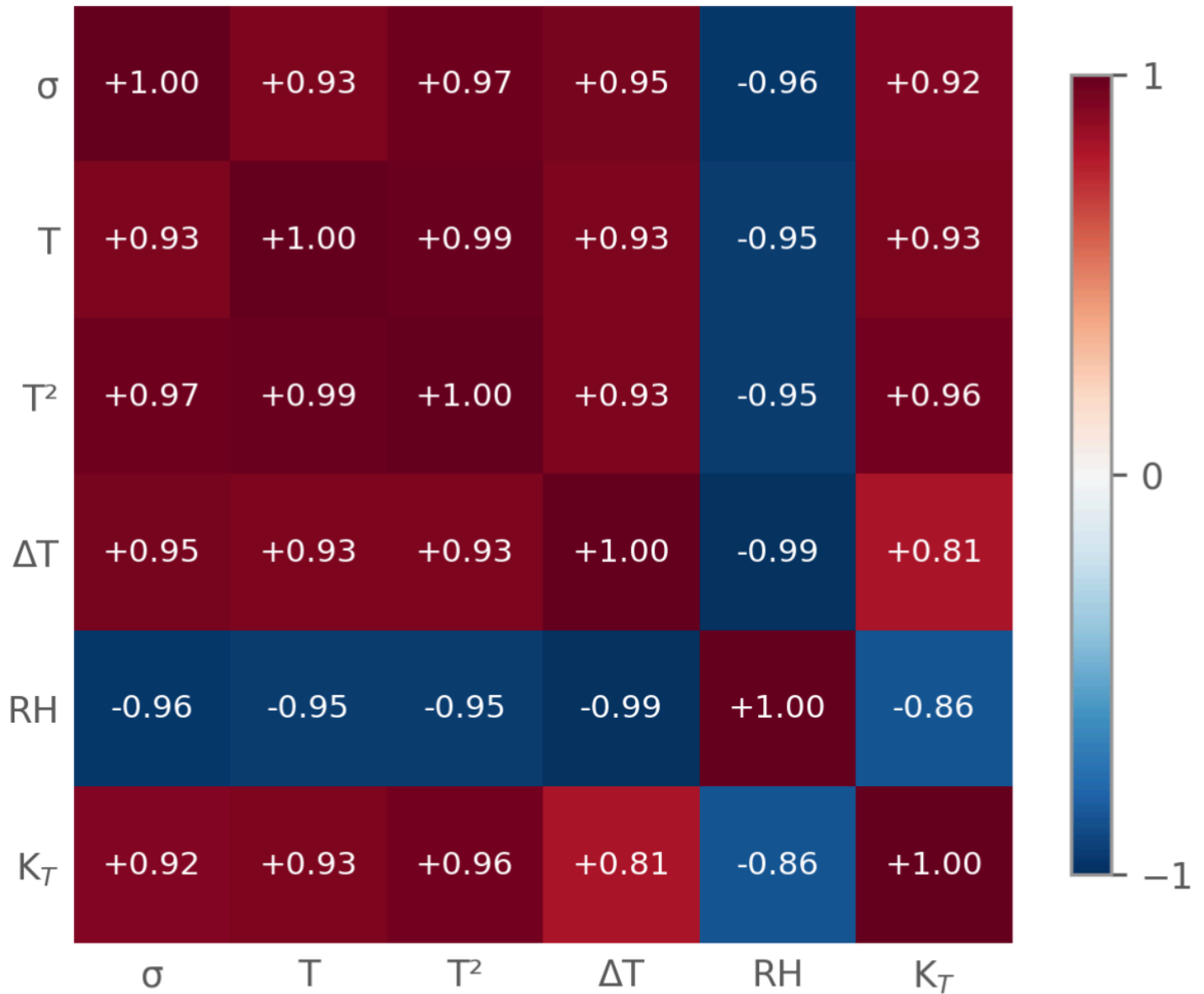


Figure 2

Correlation matrix of the predictors and of the clearness index.

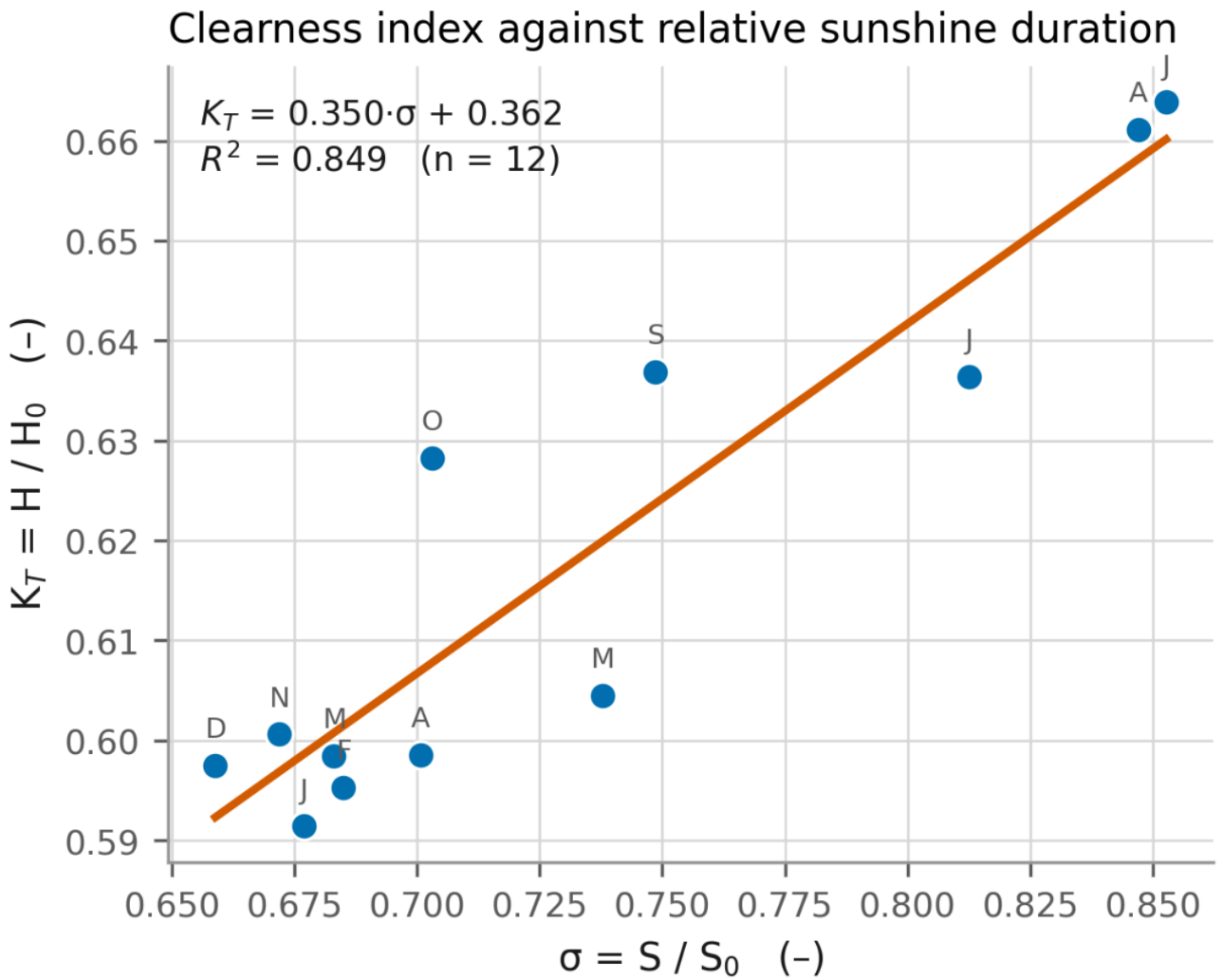


Figure 3

Clearness index against relative sunshine duration; the label of each point is the initial of the month.

All three predictors carry the same annual cycle

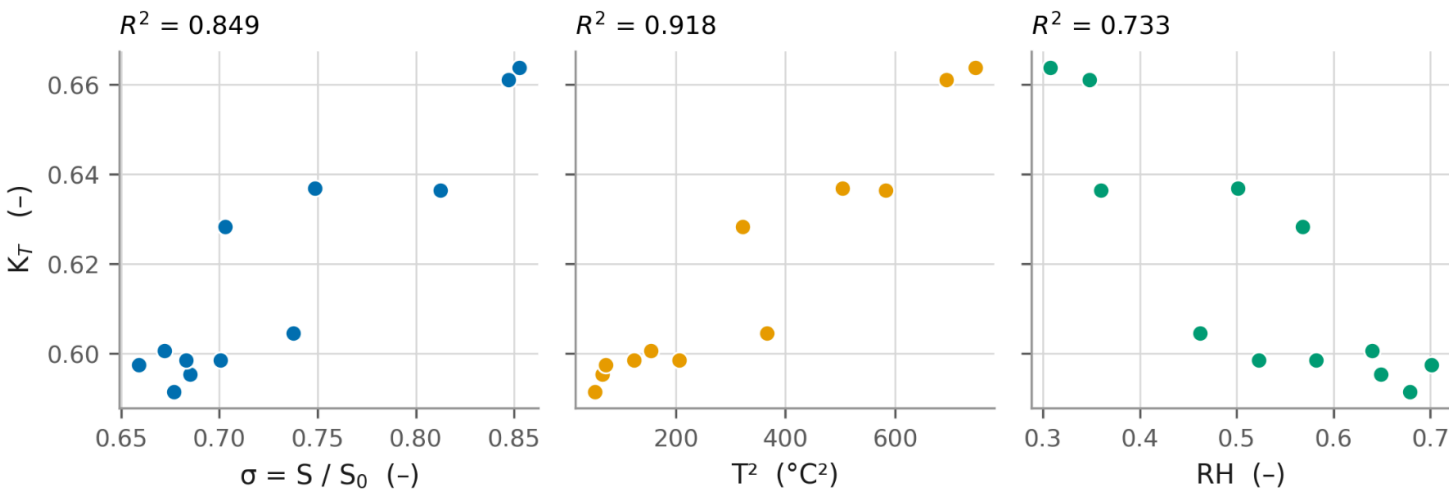


Figure 4

The clearness index against each of the three predictors taken in isolation.

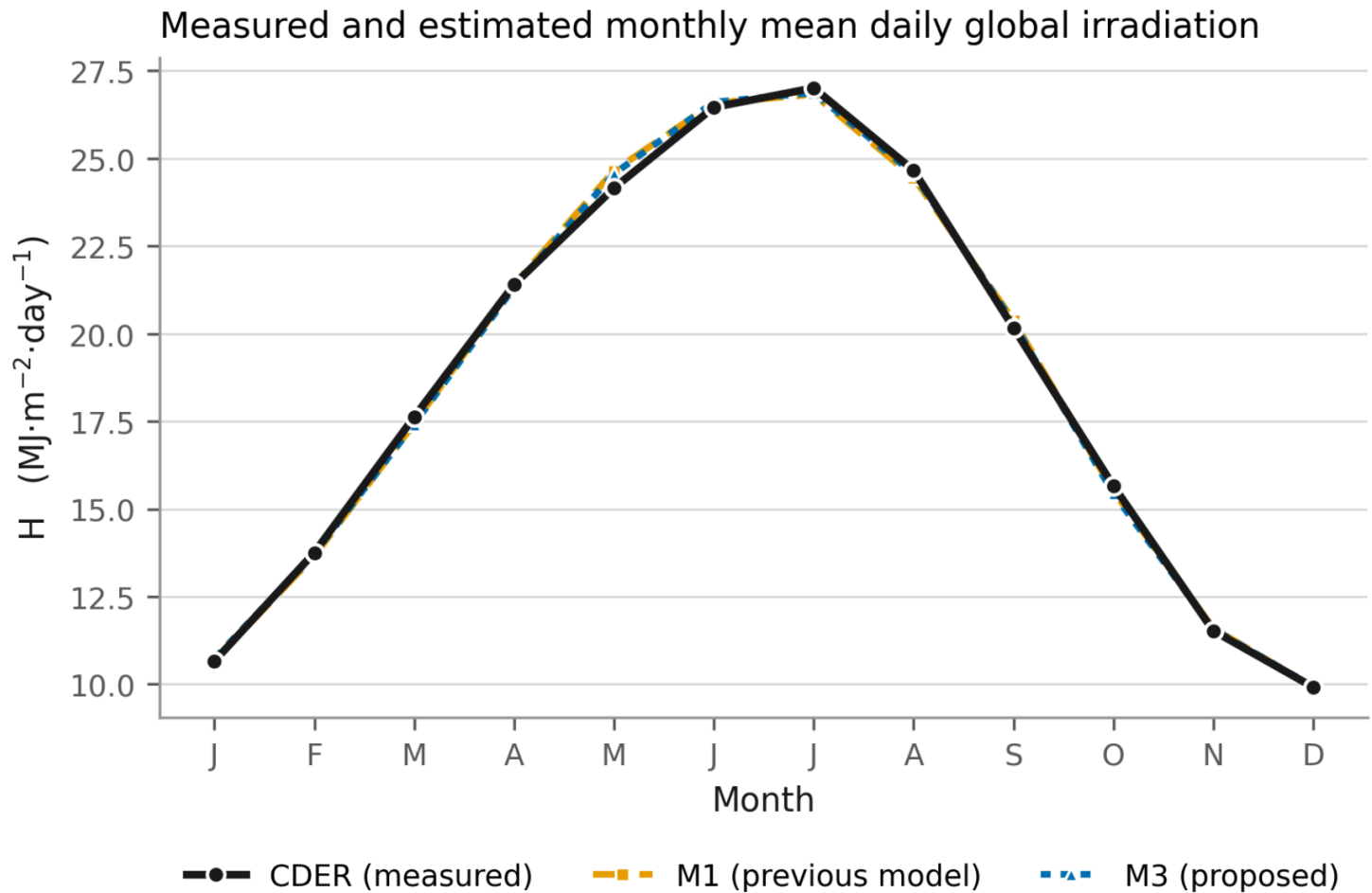


Figure 5

Measured and estimated monthly mean daily global irradiation at Tébessa.

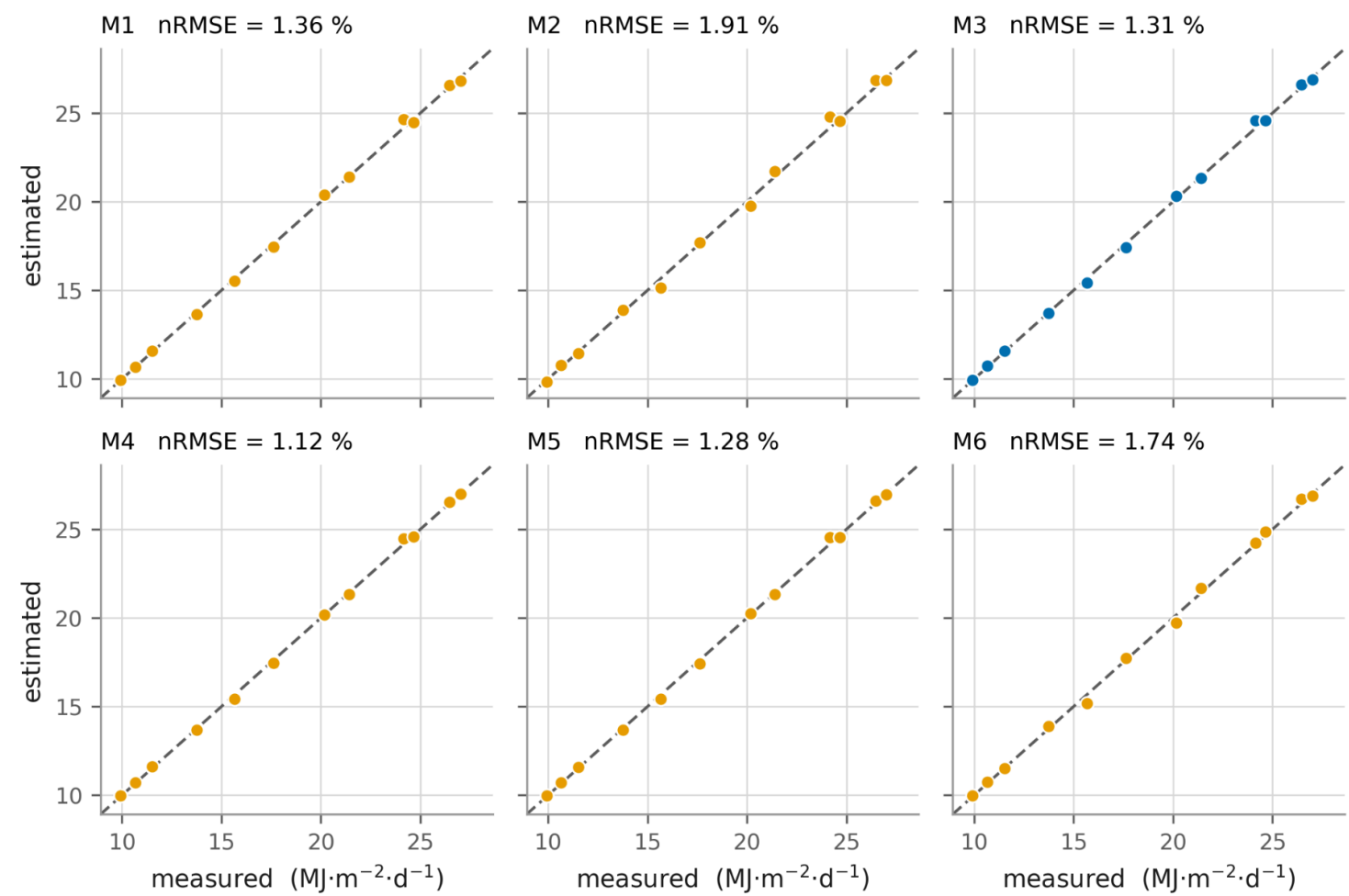


Figure 6

One-to-one plots of estimated against measured irradiation for the six candidate correlations.

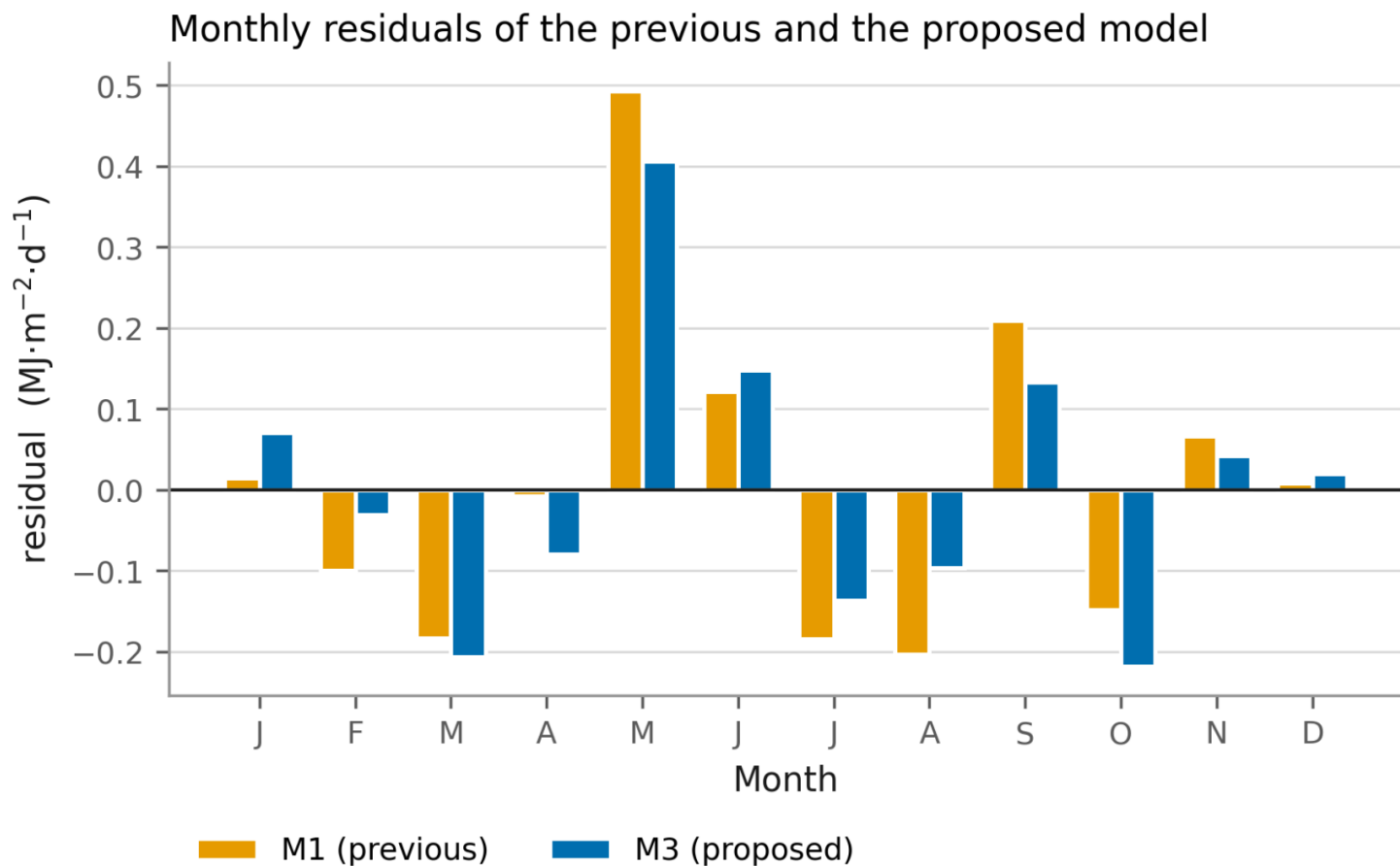


Figure 7

Monthly residuals of the previous model M1 and of the proposed model M3.

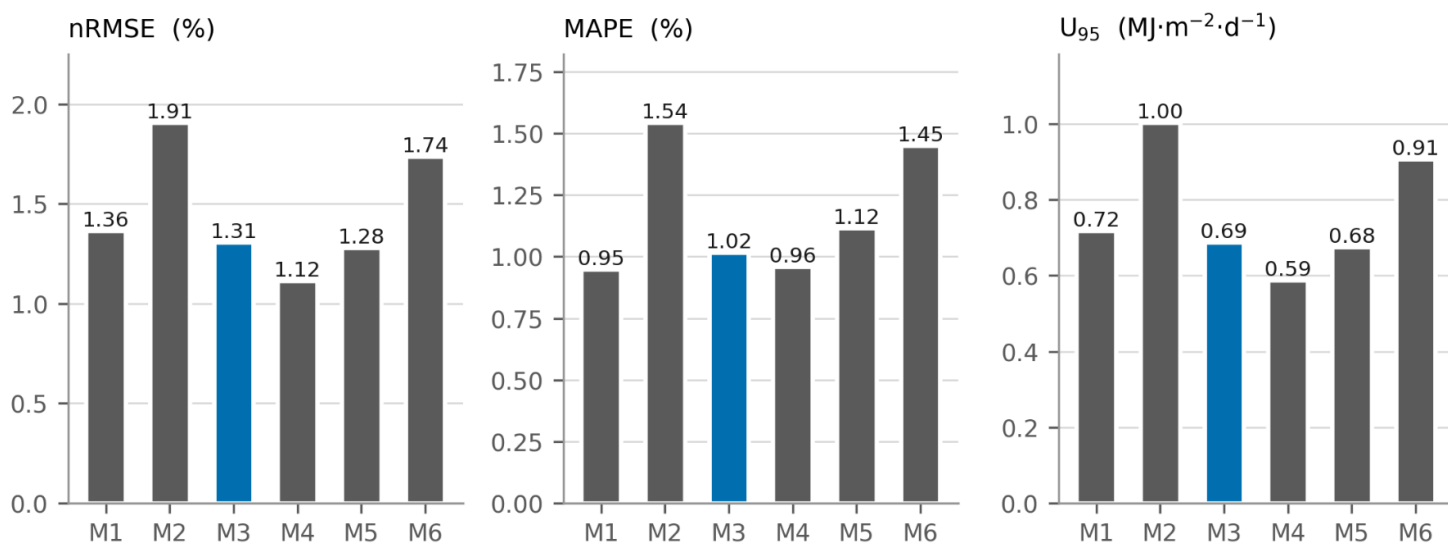


Figure 8

Normalised root mean square error, mean absolute percentage error and expanded uncertainty of the six candidate correlations.

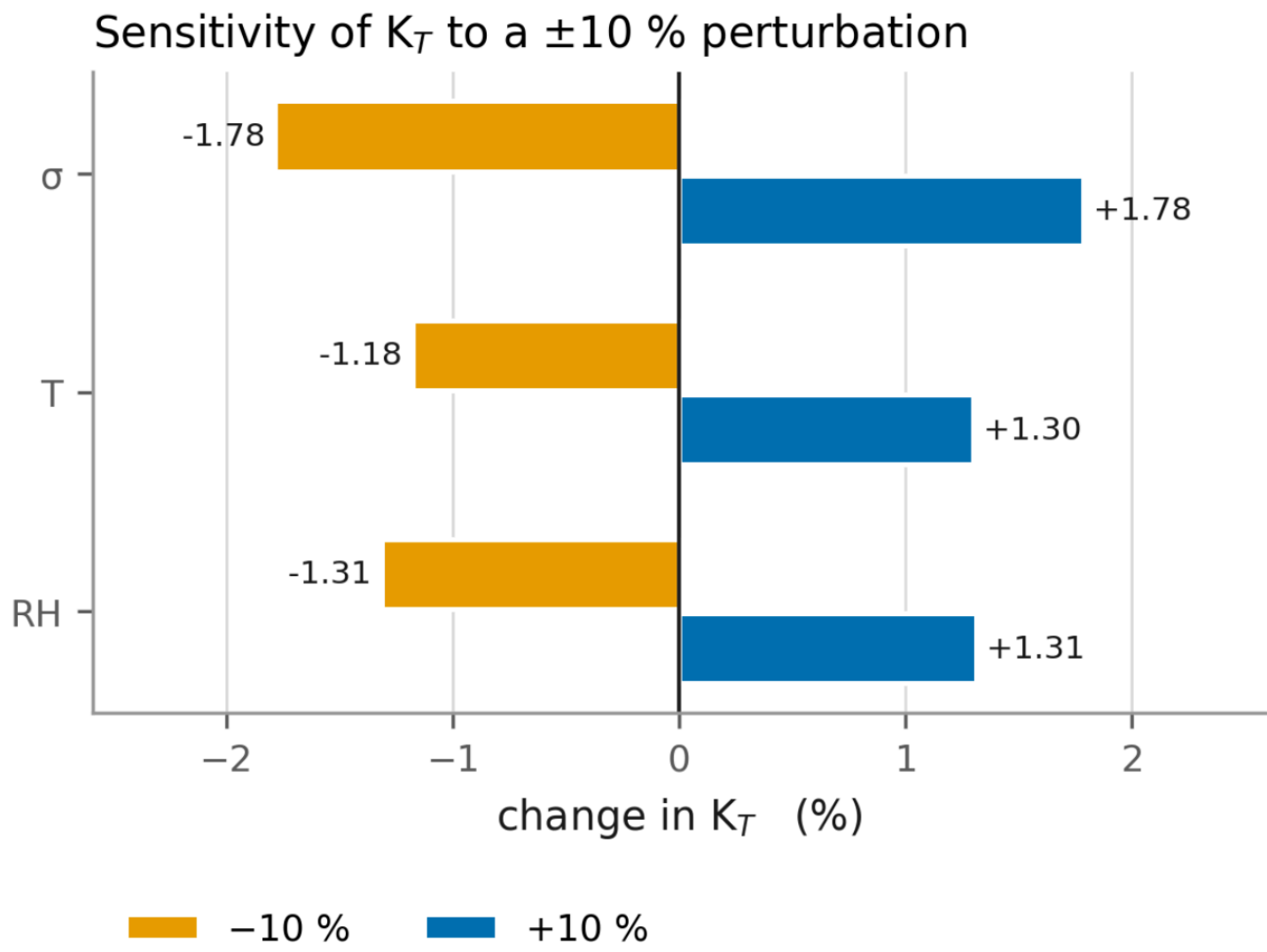


Figure 9

Sensitivity of the clearness index to a $\pm 10\%$ perturbation of each predictor of the proposed correlation.